LEVEL II

(12)

AD A058998

DDC FILE COPY

TECHNICAL REPORT PTR-1060-78-7

# IMPROVING INTUITIVE JUDGMENTS BY SUBJECTIVE SENSITIVITY ANALYSIS

DECISION RESEARCH • A BRANCH OF PERCEPTRONICS

Baruch Fischhoff
Paul Slovic
Sarah Lichtenstein

July 1978





ADVANCED DECISION TECHNOLOGY PROGRAM

CYBERNETICS TECHNOLOGY OFFICE
DEFENSE ADVANCED RESEARCH PROJECTS AGENCY
Office of Naval Research • Engineering Psychology Programs

The objective of the Advanced Decision Technology Program is to develop and transfer to users in the Department of Defense advanced management technologies for decision making. These technologies are based upon research in the areas of decision analysis, the behavioral sciences and interactive computer graphics. The program is sponsored by the Cybernetics Technology Office of the Defense Advanced Research Projects Agency and technical progress is monitored by the Office of Naval Research – Engineering Psychology Programs. Participants in the program are:

**Decisions and Designs, Incorporated**
**Harvard University**
**Perceptronics, Incorporated**
**Stanford Research Institute**
**Stanford University**
**The University of Southern California**

Inquiries and comments with regard to this report should be addressed to:

**Dr. Martin A. Tolcott**
Director, Engineering Psychology Programs
Office of Naval Research
800 North Quincy Street
Arlington, Virginia 22217

or

**Dr. Stephen J. Andriole**
Cybernetics Technology Office
Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, Virginia 22209

TECHNICAL REPORT PTR-1060-78-7

# IMPROVING INTUITIVE JUDGMENTS BY SUBJECTIVE SENSITIVITY ANALYSIS

by

Baruch Fischhoff, Paul Slovic and Sarah Lichtenstein

Sponsored by

Defense Advanced Research Projects Agency
ARPA Order No. 3469

Under Subcontract from
Decisions and Designs, Incorporated

July 1978



**DECISION RESEARCH**
**A BRANCH OF PERCEPTRONICS**
1201 Oak Street
Eugene, Oregon 97401
(503) 485-2400

## SUMMARY

### Overview

A series of three experiments investigated the feasibility of a technique designed to reduce the potency of several severe judgmental biases. Modest success was demonstrated with most participants in the studies. The technique is very simple and is applicable to a variety of tasks that occur in many decision-making situations.

### Background and Approach

A disturbing result in many earlier studies of judgment and decision making is that people tend to ignore various kinds of important information when making inferences. These included information regarding the validity and reliability of the data upon which they base their judgments. In decision-making contexts, these biases can lead to recruiting the wrong information and reaching erroneous conclusions. They seem to be quite robust and so far have resisted attempts to eliminate them.

These experimental studies have all mimicked the typical "real-life" decision-making setting by providing participants with one set of information including the critical piece of information, say, a measure of the reliability of the remaining information. Different groups of participants would be given sets of information differing only in the value of that critical datum. For example, one group might be told that the information came from a reliable source while one was told that it came from an unreliable source. The similarity of inferences made by the two groups was taken as an indication that varying that piece of information had little effect.

The debiasing technique tested here used the simple device of having participants consider how they would make inferences if provided with several values of the critical piece of information. For example, they might be asked how they make their judgments if told the information was highly reliable and if told that it was highly unreliable. In effect, they were asked to perform a sensitivity analysis on their own judgments. Such subjective sensitivity analyses were tried with three kinds of information: base-rate information (telling what was the typical occurrence in a particular situation), validity information (describing the predictive power of the remaining information), and sample size (indicating the stability of the observed information).

## Findings and Implications

For base-rate and validity information, roughly 2/3 of all participants changed their judgments as the value of the information changed. For the vast majority, these changes were in the appropriate direction, although not necessarily as large as they should be. Thus, once that information was brought to their attention, they demonstrated a knowledge of its inferential meaning not shown in previous studies. No such sensitivity was demonstrated with variations on sample size information.

Where effective, this technique appears to have potential usefulness as a debiasing procedure. It is readily applied to any kind of information and serves to improve the decision makers' intuitive feel for the presented information. Further work is needed, however, to understand why it did not work with all kinds of information and why changes in the appropriate directions tended to be too small.

TABLE OF CONTENTS

LIST OF FIGURES

## LIST OF TABLES

ACKNOWLEDGMENT

We would like to thank Nancy Collins, Barbara Combs, Michael Enbar, Gerry Hanson and Peggy Roecker for their help in conducting this study. The work was supported by the Advanced Research Projects Agency of the Department of Defense and was monitored by Office of Naval Research under Contract N00014-78-C-0100 (ARPA Order No. 3469) under Subcontract 78-030-0714 from Decisions and Designs, Inc. to Perceptronics, Inc.

# 1. INTRODUCTION

A striking conclusion from recent studies of probabilistic thinking is that people are oblivious to several kinds of information that play a major role in normative models of inference. These include information regarding sample size (Tversky & Kahneman, 1971), predictive validity (Kahneman & Tversky, 1973) and base rates (Lyon & Slovic, 1976). According to Kahneman and Tversky (1972), "the notion that sampling variance decreases in proportion to sample size is apparently not part of [people's] repertoire of intuitions" (p. 444); according to Tversky and Kahneman (1974), "subjects show little or no regard for consideration of predictability" (p. 1126); according to Lyon and Slovic (1976), "subjects' responses were determined predominantly by the specific evidence; the prior probabilities were neglected, causing the judgments to deviate markedly from the normative response" (p. 287).

In the typical experiment, subjects were given one story problem presenting several pieces of information. Failure to attend to one kind of information was demonstrated by showing similar responses in groups of subjects whose story problems differed only in the value given to that kind of information. For example, Lyon and Slovic (1976) had people assess the probability that a light bulb identified as defective by an imperfect scanner was in fact defective. A group of subjects told that a small proportion of bulbs were defective responded similarly to a group told that a large proportion were defective. That is to say, base-rate information had no apparent effect on judgments.

One interpretation of such results is that people believe that such information was not relevant to the story problem. A second interpretation is that they realize the importance of the information, but lose it in the cognitive shuffle of combining various pieces of information to obtain a summary judgment. If the second interpretation is correct, the impact of otherwise ignored information might be increased by simply highlighting its salience. Ajzen (1977) and Bar-Hillel (1977) increased the salience of base-rate information by giving it a causal relation to the other information. In these contexts, it was no longer neglected.

The present studies adopted one strategy to highlight the importance of information not used in earlier studies: having subjects consider how alternative values of the datum in question would affect their judgment. In effect, the between-subject designs of earlier experiments were converted to within-subject designs, each subject considering values previously considered by separate groups. If subjects do respond differently when confronted with different values, one may surmise either (a) that they knew all along what that datum meant and only needed help to attend to it or (b) that they never knew or had thought about its meaning, but once posed the question, were able to figure out what it meant. In either case, the earlier conclusion that "people ignore . . . information" would have to be qualified.

Asking subjects to make the same judgment several times while varying the value imputed to one variable contains an implicit demand that they change their responses somehow. Refusal to change makes a strong statement regarding the irrelevance of the varied piece of information. Even if subjects shift their responses, they need not do so in the

proper direction (or with the proper magnitude), unless they have some understanding of the meaning of the varied datum (or could figure it out on the spot).

Whatever their theoretical importance, such within-subject manipulations might have applied implications. Decision analysts, the purveyors of formal decision-making techniques, test the robustness of their recommendations by reworking the decision problem with different values of key variables (e.g., probability of success, expected gain). Such repeated analyses are called "sensitivity analyses," since they test the sensitivity of the final decision to variations in the inputs. The within-subject designs used here essentially force subjects to perform a sensitivity analysis on their own judgments. If this procedure proves effective, one might offer judges the following general advice: "Let the value of each piece of information you are given vary through the range of reasonable values. . Consider how you would make your summary judgment given each of these (hypothetical) values. Then you will have a better appraisal of the meaning of the values you did receive."

## 2. EXPERIMENT 1--BASE-RATE INFORMATION

Earlier work has convincingly shown that people often ignore base-rate information when given individuating information (Kahneman & Tversky, 1972; Lyon & Slovic, 1976; Nisbett & Borgida, 1975). The only exceptions seem to be when base-rate information is given causal relevance (Ajzen, 1977; Bar-Hillel, 1977). Some base-rate information, however, has only diagnostic relevance. Even though it should affect judgments regarding the target (judged) event, it is not causally linked to that event. The present study attempted to induce subjects to see the relevance of diagnostic, but non-causal, information by confronting each subject with several values of that information.

One reason for modest optimism regarding this manipulation is a result from Fischhoff (1977). There subjects were asked to make causal attributions about pairs of events differing only in consensus information (describing what most people did in that situation). One version of the event asserted that almost everybody did the act in question; the second version asserted that almost nobody did it. When groups of subjects were given but one of the two versions of each vignette, they made almost identical attributions. That is, consensus information (which is the attributional equivalent of base-rate information) was ignored, or at least made no appreciable impact on their judgments. Other groups of subjects were asked to consider both possible values of the consensus information. They were asked, "How would you make your attributions if you learned that almost everyone acted this way?" and "How would you . . . if . . . almost no one acted this way?" Here, consensus information had a

substantial impact. People thought that variations in consensus information would drastically change their attributions. Consensus information in those settings may have had some causal relevance: The fact that everyone else acted one way or another might have been interpreted as implying a social norm directly affecting the actor's decision of how to behave. By and large, however, the relevance of consensus information was non-causal; it was only a diagnostic sign describing how most people were affected by the constraints of a situation.

Whether such forced sensitivity analysis will generally induce people to appreciate the significance of non-causal base-rate information was studied using two problems that have proven most impervious to previous debiasing attempts (Kahneman & Tversky, 1973; Lyon & Slovic, 1976). One was the "cab problem," the basic version of which reads as follows:

> Two cab companies, the Blue and the Green, operate in a given city. Eighty-five percent of the cabs in the city are Blue; the remaining 15% are Green. A cab was involved in a hit-and-run accident at night. A witness identified the cab as a Green cab.
>
> The court tested the witness' ability to distinguish a Blue cab from a Green cab at night by presenting to him film sequences, half of which depicted Blue cabs, and half depicting Green cabs. He was able to make correct identification in 8 out of 10 tries. He made one error on each color of cab.
>
> What do you think is the probability (expressed as a percentage) that the cab involved in this accident was Green?

Here the base rate of Green cabs is 15% and the correct value of P (Green|witness says green) = $P(G|g) = .41$. Lyon and Slovic found a median response of .80 for this version, .75 for a version with the order of the individuating and base-rate information reversed, and .20 (instead of .59) for the same problem with the last word changed to Blue. Other variants of this problem and analogous ones showed equally erroneous judgments.

The second problem was the "light-bulb problem" developed by Lyon and Slovic (1976). It read:

> A light bulb factory uses a scanning device which is supposed to put a mark on each defective bulb it spots in the assembly line. Eighty-five percent of the ligh bulbs on the line are OK; the remaining 15% are defective.
>
> The scanning device is known to be accurate in 80% of the decisions, regardless of whether the bulb is actually OK or actually defective. That is, when a bulb is good, the scanner correctly identifies it as good 80% of the time. When a bulb is defective, the scanner correctly marks it as defective 80% of the time.
>
> Suppose someone selects one of the light bulbs from the line at random and gives it to the scanner. The scanner marks this bulb as defective.
>
> What do you think is the probability (expressed as a percentage) that this bulb is really defective?

Since the base rate of defectives, $P(D)$, is 15% and the diagnosticity of the scanner is like that of the witness above, the correct answer here, too, is $.41 = P(D|d) = P$ (Defective bulb|scanner says defective).

In the present experiment, subjects were asked to consider several versions of either the cab or the light bulb problem that differed only in the base rate provided. Later in the experimental session, they were given the basic version of the other (light bulb or cab) problem as a generalization test. If varying the base rate does improve judgments on one task, perhaps it will heighten sensitivity to base-rate information on an analogous problem.

## Method

The design was a 2 x 2 x 2 x 2 factorial. Roughly half of all subjects made a series of judgments on versions of the cab problem differing only in the value of base-rate information, P(G); later they considered the basic version of the light bulb problem. The remaining subjects first judged different versions of the light bulb problem and later the basic version of the cab problem.

The second factor was the extremity of the base rates considered. Subjects either considered base rates, P(G) or P(D), of .02, .15 and .98 or of .15 and .85. Although .15 and .85 are quite discrepant values, it was thought that consideration of a situation in which the apparently observed event (a green cab or a defective bulb) was extremely typical (.98) or atypical (.02) might be needed to signal the base rate's importance. The third factor was whether the highest (.98 or .85) or lowest (.02 or .15) base rate was considered first. The fourth factor was whether or not subjects made an initial judgment before being given any base-rate information. Such a judgment could serve as an anchor for future judgments, making them less responsive to subsequent changes in base rates.

On the other hand, such a judgment could afford an additional opportunity to reflect on the significance of base-rate information (or its absence) and increase sensitivity to changes. The different experimental groups are identified in the bottom section of Table 1.

The various base rates were introduced with appropriately worded phrases, "let us say that it is now revealed that only 2% of the cabs in the city are Green," "On the other hand, let us say that . . . " and "in fact, only 15% . . . ."

Several unrelated tasks separated the problem in which subjects considered several base rates and the generalization problem for which they considered only the basic version (with the base rate of .15). No explicit mention was made of any connection between the tasks.

Subjects were 346 individuals who responded to an advertisement in the University of Oregon student newspaper. The number of subjects in each group is also presented at the bottom of Table 1. The present tasks were self-paced and embedded in a 90-minute experimental session involving a variety of unrelated judgment tasks.

## Results

The distributions of responses were highly skewed, with a portion of subjects (16% over all groups) responding .8 every time they were asked to assess $P(G|g)$ or $P(D|d)$. As a result, both means and medians are presented for most analyses.

As the top part of Table 1 shows, the proportion of subjects who always said .8 varied from 0% to 33% over the different groups. There appears to be no consistent pattern to these percentages. For example, there is little relation between the percentages for corresponding cab-first and light bulb-first groups. The top section of Table 2 collapses these proportions over the various factors in our design. Other than a somewhat higher proportion with the groups who considered varying base rates with the light bulb problem, there is no obvious pattern. We cannot reject the possibility that these proportions reflect a random distribution over the groups of subjects who refuse to attend to base rates. Since the percentage of resolute .8 responders could have substantial impact on group results, later analyses were conducted both with and without these subjects. No different conclusions were reached.

Always responding .8 is one heuristic device for dealing with base-rate information: ignore it. An alternative, and equally extreme, heuristic is to ignore the diagnostic information and always respond with the base rate. The second sections of Tables 1 and 2 show how many subjects adopted this strategy. About 10% of all subjects ignored the individuating information, the overwhelming majority of whom did so in response to the light bulb problem. The reasons for this discrepancy are unclear.

An alternative strategy, and a more normatively appropriate one, is to combine base-rate and individuating information. The third sections of Tables 1 and 2 show the proportion of subjects whose assessments of $P(G|g)$ or $P(D|d)$ were ordered according to the base rates. Almost two-thirds

Table 1

Percentages of Subjects with Different Response Patterns (by Group)

| Problem Considered First | | low base rate first (2, 15, 98) | (15, 85) | high base rate first (2, 15, 98) | (15, 85) |
|---|---|---|---|---|---|
| | | **All Responses = .80** | | | |
| Cab | anchor | 0 | 21 | 3 | 7 |
| | no anchor | 33 | 11 | 24 | 5 |
| light bulb | anchor | 16 | 10 | 26 | 30 |
| | no anchor | 33 | 13 | 18 | 21 |
| | | **Responses = Base Rate** | | | |
| Cab | anchor | 4 | 0 | 3 | 0 |
| | no anchor | 0 | 0 | 24 | 5 |
| light bulb | anchor | 42 | 10 | 26 | 15 |
| | no anchor | 11 | 13 | 12 | 21 |
| | | **Responses Ordered According to Base Rates** | | | |
| Cab | anchor | 70 | 48 | 66 | 83 |
| | no anchor | 38 | 68 | 62 | 74 |
| light bulb | anchor | 68 | 75 | 63 | 50 |
| | no anchor | 61 | 56 | 71 | 58 |
| | | **Group Code and Number of Subjects** | | | |
| Cab | anchor | Y (27) | X (33) | Z (29) | W (29) |
| | no anchor | Y'(21) | X'(19) | Z'(21) | W'(19) |
| light bulb | anchor | Y (19) | X (20) | Z (19) | W (20) |
| | no anchor | Y'(18) | X'(16) | Z'(17) | W'(19) |

Table 2

Percentages of Subjects with Different Response Patterns
(by Factor)

| Problem Considered First | Anchor: With (WXYZ) | Anchor: Without (W'X'Y'Z') | Values: 2,15,98 (YZY'Z') | Values: 15,85 (WXW'X') | Order: low 1st (XYX'Y') | Order: high 1st (WZW'Z') | All |
|---|---|---|---|---|---|---|---|
| **All Responses = .80** | | | | | | | |
| cab | 8* | 19 | 13 | 12 | 16 | 9* | 13 |
| light bulb | 21 | 21 | 23 | 19 | 18 | 24 | 21 |
| both | 13 | 20 | 18 | 15 | 17 | 16 | 16 |
| **Responses = Base Rate** | | | | | | | |
| cab | 2*** | 8 | 7** | 1** | 1*** | 7 | 4*** |
| light bulb | 23 | 14 | 23 | 15 | 19 | 19 | 19 |
| both | 10 | 11 | 14 | 7 | 9 | 12 | 10 |
| **Responses Ordered According to Base Rates** | | | | | | | |
| cab | 66 | 60 | 59 | 67 | 56 | 71 | 64 |
| light bulb | 64 | 61 | 66 | 60 | 66 | 60 | 63 |
| both | 65 | 61 | 63 | 64 | 60 | 66 | 63 |

Note: Within each column asterisks indicate paired entries that statistically different (* $p < .05$; ** $p < .01$; *** $p < .001$; two-tailed). No entries in adjacent rows were significantly different.

of subjects were sensitive to the base rates in this sense, a remarkably high percentage considering the usual conclusion that subjects "ignore" base-rate information.

The extent to which subjects attend to base-rate information can only be assessed by considering the numerical value of their responses (and not just their order). Table 3 shows mean and median results for all groups on the problem for which base rates were manipulated. Clearly, these probabilities reflected the base rates. Subjects shown a base rate of .98 gave the highest values to $P(G|g)$ or $P(D|d)$, while those shown .02 gave the lowest values. A striking contrast to this sensitivity is provided by one of Lyon and Slovic's groups, which produced a median of .8 even with a base rate of .01. However, the present values were not optimal; they tended to lie between the optimal value and .8, the accuracy rate for the witness or the scanner, which appeared to serve as a powerful anchor.

Both the cab and light bulb problems have been most heavily studied with the base rate equal to .15. For that reason, .15 was the one base rate appearing in all conditions. The overall median response with this base rate was .64 for the cab problem and .53 for the light bulb problem. In past studies, the median response has typically been .8. This modest difference between the cab and light bulb problems may be attributed to the higher proportion of subjects who always responded with the base rate in the light bulb problem.

One measure of the effectiveness of the manipulations, shown in Table 4, is the proportion of subjects who assigned a value to $P(G|g)$ or $P(D|d)$ lower than the median (.8)

Table 3

Probability Assessments for Manipulation Problem

| Group | Order | Base Rate Given: Mean None | 98 | 85 | 15 | 2 | Median None | 98 | 85 | 15 | 2 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Cab Problem | | | | | | | | | | | |
| W | N-85-15 | 76 | -- | 88 | 61 | -- | 80 | -- | 90 | 65 | -- |
| X | N-15-85 | 67 | -- | 73 | 58 | -- | 80 | -- | 80 | 65 | -- |
| Y | N-2-98-15 | 72 | 82 | -- | 52 | 42 | 80 | 85 | -- | 60 | 50 |
| Z | N-98-2-15 | 70 | 92 | -- | 53 | 41 | 80 | 90 | -- | 60 | 40 |
| W' | 85-15 | -- | -- | 79 | 66 | -- | -- | -- | 80 | 75 | -- |
| X' | 15-85 | -- | -- | 73 | 59 | -- | -- | -- | 80 | 60 | -- |
| Y' | 2-98-15 | -- | 83 | -- | 64 | 55 | -- | 80 | 80 | 70 | 70 |
| Z' | 98-2-15 | -- | 87 | -- | 48 | 39 | -- | 90 | -- | 50 | 20 |
| | All | 71 | 86 | 78 | 57 | 44 | 80 | 85 | 80 | 64 | 50 |
| Light bulb problem | | | | | | | | | | | |
| W | N-85-15 | 76 | -- | 78 | 55 | -- | 80 | -- | 82 | 68 | -- |
| X | N-15-85 | 72 | -- | 78 | 57 | -- | 80 | -- | 85 | 60 | -- |
| Y | N-2-98-15 | 69 | 89 | -- | 39 | 29 | 80 | 98 | -- | 35 | 2 |
| Z | N-98-2-15 | 63 | 91 | -- | 40 | 29 | 80 | 98 | -- | 15 | 2 |
| W' | 85-15 | -- | -- | 76 | 61 | -- | -- | -- | 80 | 79 | -- |
| X' | 15-85 | -- | -- | 78 | 49 | -- | -- | -- | 83 | 50 | -- |
| Y' | 2-98-15 | -- | 85 | -- | 61 | 47 | -- | 93 | -- | 80 | 73 |
| Z' | 98-2-15 | -- | 90 | -- | 50 | 32 | -- | 90 | -- | 55 | 10 |
| | All | 70 | 89 | 77 | 51 | 34 | 80 | 95 | 85 | 53 | 10 |
| OPTIMAL | | 80 | 99.5 | 95.8 | 41.4 | 7.5 | 80 | 99.5 | 95.8 | 41.4 | 7.5 |

observed by Lyon and Slovic: over 60% of the subjects did so, while 9% assigned values higher than .8. A higher percentage of responses below .8 was observed with the anchor groups than with the no-anchor subjects (70% vs. 57%, z = 2.43), with the cab problem than with the light bulb problem (68% vs. 59%, z = 1.71), and with subjects who considered the more extreme base rates (68% vs. 60%, z = 1.51).

Tables 4 and 5 describe the results of the generalization problems. The analogy between the structure of this problem and the preceding problem with a base rate of .15 was obviously not apparent to subjects. Only 72 of 346 (21%) assigned the same probability to both problems: of these, 24 were subjects who always assigned 18. Performance on the generalization problem was nonetheless slightly better than that observed elsewhere; 180 (52%) subjects assigned values lower than the median assigned in Lyon and Slovic (.8); 117 (34%) assigned that value and 49 (11%) assigned higher values. Still, their responses were far from the optimal value of .41. There was no indication that subjects whose responses were ordered according to the different values of the base rate in the first problem were more accurate than other subjects in the generalization problem. Nor were any systematic differences in responses to the generalization problem associated with any of the factors of the manipulation problem.

Discussion

The robust finding of these studies is that most people know (or guess) the direction in which base rates should influence their judgments. Over the two problems, 65% of subjects ordered their probability assessments according to the base rates. However, they did not adjust enough. In

Table 4

Percentage of Responses $\lesseqgtr$ .8
When Base Rate = .15

| Manipulation Problem | | | Generalization Problem | |
|---|---|---|---|---|
| Cab | <.8 | 67.7 | 48.0* | Light Bulb |
| | =.8 | 22.7 | 37.3* | |
| | >.8 | 9.6 | 14.7 | |
| Light Bulb | <.8 | 58.8 | 57.4 | Cab |
| | =.8 | 32.4 | 29.1 | |
| | >.8 | 8.8 | 13.5 | |

* $p < .001$ difference within row

addition, generalization to the second problem was limited and no better among subjects who were sensitive to base-rate variation in the first problem than among subjects who were not. The most obvious explanation is that subjects just did not see the similarity of structure of the two problems.

Although training in statistics may be necessary to help subjects see structural similarity, the modest effects found with different modes of presenting base-rate information suggest technical solutions for improving the judgments of subjects who are somewhat but not sufficiently sensitive to changes. Subjects assigned lower and more optimal probabilities for the .15 base rate both (a) when they had first made an assessment with no base rate and (b) when they considered a set of extreme base rates (.98 and .02). The most accurate assessments of all were found in the groups (Y,Z) who both made a no-base-rate judgment <u>and</u> considered the extreme values. If these manipulations are generally effective, they could be incorporated as standard features in judgmental exercises. Their effectiveness might be traced either to alerting subjects to some otherwise unnoticed implication or base-rate information or to the fact that each required subjects to make one additional judgment. Making more responses might make responses more optimal by increasing their range and moving them away from the anchor (.8) provided by the diagnosticity information (see also Selvidge, 1975).

The study of subjective sensitivity analyses has two aspects. The first is discovering whether people will use a particular kind of information in making their judgments. The second is discovering in what way the information will

Table 5

Probability Assessments for Generalization Problem

| | Mean | | Median | |
|---|---|---|---|---|
| Group | All Ss | Ss Ordering Manipulation Correctly | All Ss | Ss Ordering Manipulation Correctly |
| Cab Problem first; Light Bulb Problem Generalization | | | | |
| W | 55 | 55 | 70 | 74 |
| X | 60 | 56 | 80 | 79 |
| Y | 54 | 50 | 70 | 60 |
| Z | 67 | 60 | 80 | 75 |
| W' | 70 | 71 | 80 | 80 |
| X' | 65 | 69 | 80 | 80 |
| Y' | 51 | 41 | 50 | 40 |
| Z' | 72 | 64 | 80 | 80 |
| Total | 61 | 58 | 80 | 75 |
| Light Bulb Problem First; Cab Problem Generalization | | | | |
| W | 67 | 68 | 70 | 65 |
| X | 62 | 58 | 70 | 70 |
| Y | 60 | 63 | 70 | 80 |
| Z | 64 | 60 | 75 | 75 |
| W' | 66 | 74 | 75 | 75 |
| X' | 59 | 64 | 68 | 75 |
| Y' | 60 | 55 | 76 | 60 |
| Z' | 69 | 66 | 79 | 77 |
| Total | 63 | 63 | 75 | 70 |

Note: optimal answer = 41

be used. In the present study, 84% of the subjects altered their responses as the base rate changed; of those, three quarters ordered their assessments according to the base rates. However, few subjects (other than those who relied exclusively on the base rates) made adjustments from the anchor value of .8 that were even close to the adjustment needed. Thus, the failure of subjects to attend to base rates in earlier studies should be interpreted as just that, failure to attend, rather than inability to respond to base rates.

## 3. EXPERIMENT 2--VALIDITY INFORMATION

In the earlier (between-subject) work on the light bulb and cab problems, the modal response (.8) reflected complete reliance on the validity of the information source (that is, the accuracy of the scanner or witness). Other research by Kahneman and Tversky (1973) has shown that in other circumstances validity information has no impact on judgments. In one of their demonstrations, subjects were asked to estimate the Grade Point Average (GPA) associated with each of a series of percentile scores. The percentile scores came from one of three sources: the distribution of GPA's a test of mental concentration, and a test measuring sense of humor. These sources were described as having high, medium and low validity, respectively, as predictors of GPA. They had each of three groups of subjects consider a set of 11 percentile scores (5, 15, 25, . . . , 85, 95) described as coming from one of these three sources. If subjects were sensitive to validity information, then predictions based on less valid scores should be more regressive (less extreme) than those based on more valid sources. Kahneman and Tversky (1973) found, however, no difference between predictions of GPA based on GPA and mental concentration percentile scores; that is, they showed the same range, mean and slope. Predictions made from information about sense of humor were slightly regressed (and slightly elevated) suggesting very modest sensitivity to the implications of their minimal validity.

The present experiment follows the logic of the previous one. The between-group experiment is converted to within-group form to see if subjects are sensitive to systematic variations in validity information.

## Method

Design. Subjects predicted the GPA associated with the 5th, 15th, 25th, . . . , and 95th percentiles of three distributions of scores. One set of percentiles was described as coming from the distribution of GPA's, one came from scores on a test of mental concentration described as having a moderate correlation with GPA; and one came from a measure of sense of humor described as having a low but positive correlation with GPA. Half the subjects received the sets in the order GPA, mental concentration and sense of humor; half received them in the reverse order.

These 11 percentile scores (5, 15, 25, . . . , 85, 95) were those used by Kahneman and Tversky (1973). In the present study, a second group of subjects received three sets of percentile scores with only three values (10, 50 and 90). If subjects are unwilling to assign identical GPA's to different percentiles, use of 11 scores places a minimum range on the predicted GPA's (e.g., a range of 1.0 if the minimum acceptable difference is 0.1; 2.5 if the minimum is 0.25), reducing the possibilities for regression. This difficulty would not be encountered with only three percentiles; one could assign different GPA's to different percentile scores (e.g., 1.9 to 10, 2.0 to 50 and 2.1 to 90) and still have a small range of responses. As with the 11-percentile groups, half of the 3-percentile subjects considered GPA, mental concentration and sense of humor in that order; for the remainder the order was reversed.

Instructions. Kahneman and Tversky's (1973, pp. 245-6) instructions were used verbatim with the following additional general introduction:

In this task you will be given a percentile score for each of several hypothetical students and asked to predict their grade point average at the end of their first year in college. A percentile score of 65 means that that student scored higher than 65% of all other first-year students; a percentile score of 5 means scoring higher than 5% of all other first-year students, and so on.

Subjects. Eighty-six individuals were recruited as before, roughly equal numbers serving in each condition. One group (N = 42) received the 11-percentile forms; the other group (N = 44) received the 3-percentile forms.

## Results

11 percentiles. Figure 1a shows the mean responses by all subjects making the 11-percentile judgments (combining both orders of presentation).[1] The sense of humor judgments were markedly regressed relative to the GPA and mental concentration judgments, with mental concentration judgments somewhere in between GPA and sense of humor. As indicated by Table 6, the range of responses decreased by roughly 1/3 over the conditions. These differences contrast with the virtual identity of Kahneman and Tversky's (1973) mental concentration and GPA groups and slight regression with the sense of humor group. Table 7 shows within-subject comparisons of the ranges of responses given with different sets of scores. In most cases (57%), subjects gave a larger range with the more valid score.

Table 6

Mean Range[a]

| | Test | | | |
|---|---|---|---|---|
| | GPA | Mental Concentration | Sense of Humor | N |
| **All Subjects** | | | | |
| 3 percentiles | | | | |
| GPA first | 1.99 | 1.78 | 0.85 | 21 |
| Sense of humor first | 2.01 | 1.63 | 1.32 | 23 |
| 11 percentiles | | | | |
| GPA first | 3.10 | 2.87 | 1.40 | 21 |
| Sense of humor first | 2.83 | 2.40 | 2.00 | 21 |
| Kahneman-Tversky (1973) | 2.76 | 2.78 | 2.40 | |
| **Monotonic Subjects** | | | | |
| 3 percentiles | | | | |
| GPA first | 2.19 (20)[b] | 1.94 (20) | 1.36 (17) | 21 |
| Sense of humor first | 2.18 (22) | 1.78 (22) | 1.54 (19) | 23 |
| 11 percentiles | | | | |
| GPA first | 3.10 (21) | 2.87 (21) | 2.41 (12) | 21 |
| Sense of humor first | 2.83 (21) | 2.51 (17) | 2.44 (15) | 21 |

[a] Difference between GPA associated with the highest and lowest percentiles.

[b] Number of monotonic subjects in parentheses.

Table 7

Order of Ranges
Monotonic Subjects

| | GPA - MC[a] | | | GPA-SofH | | | MC-SofH | | | Total | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | + | - | = | + | - | = | + | - | = | + | - | = |
| 3 percentiles | | | | | | | | | | | | |
| GPA first | 9 | 5 | 6 | 12 | 1 | 4 | 11 | 3 | 3 | 32 | 9 | 13 |
| Sense of humor first | 13 | 7 | 2 | 13 | 2 | 4 | 11 | 6 | 3 | 37 | 15 | 8 |
| 11 percentiles | | | | | | | | | | | | |
| GPA first | 13 | 5 | 3 | 9 | 1 | 2 | 8 | 2 | 2 | 30 | 8 | 7 |
| Sense of humor first | 8 | 3 | 6 | 7 | 6 | 2 | 5 | 7 | 3 | 20 | 16 | 11 |
| Total | 43 | 20 | 17 | 41 | 10 | 12 | 35 | 18 | 10 | 117 | 48 | 39 |

[a] Greater range with the more valid score indicated by +; lesser range indicated by -; equal range indicated by =.

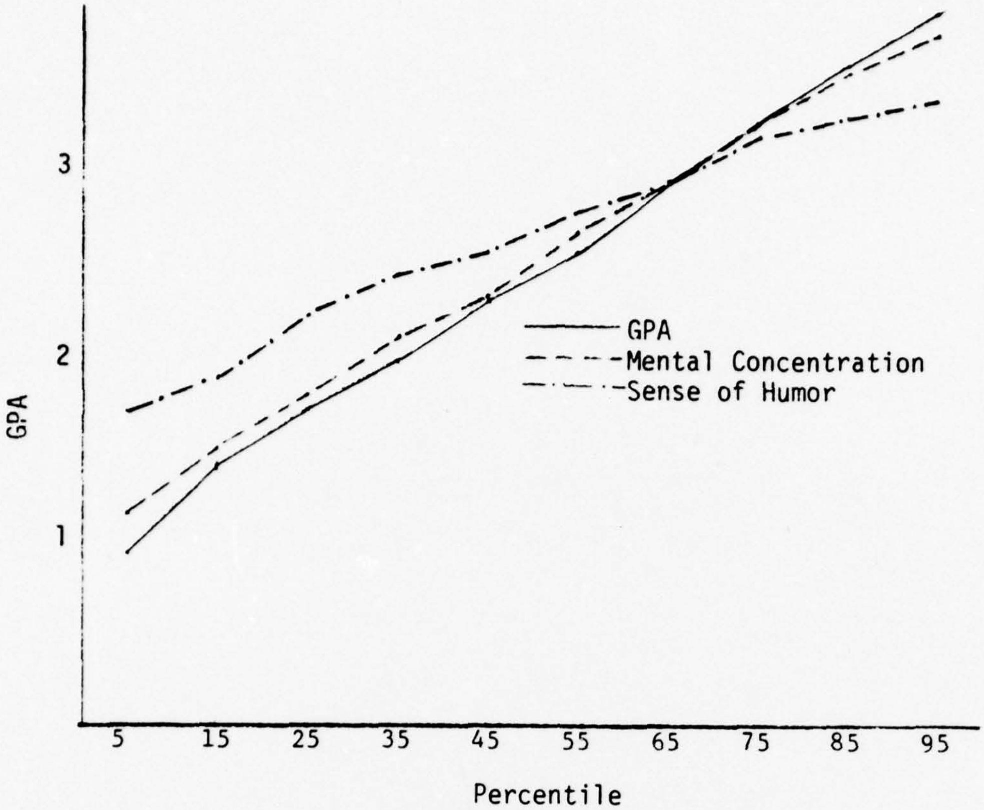


Figure 1a

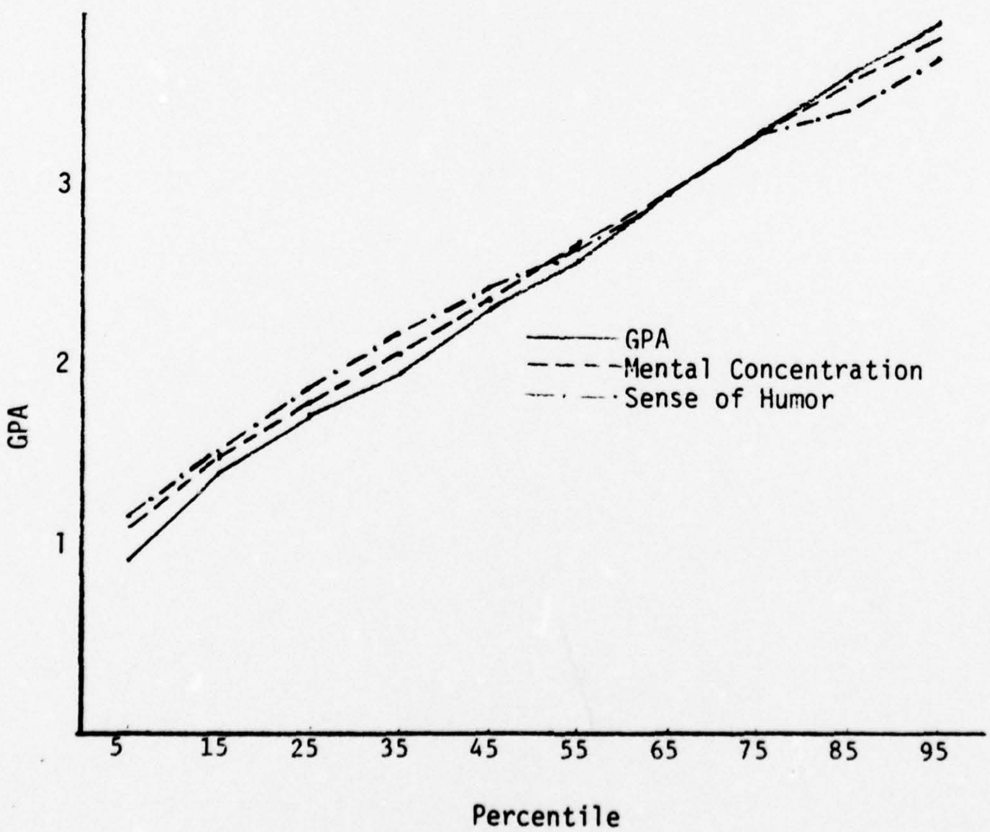


Figure 1b

Predictions of grade point average from 11 percentile scores on 11 variables. (a) All subjects (b) All subjects providing monotonic responses

Closer examination of the data revealed one problematic aspect. Some subjects exhibited a non-monotonic relationship between percentile scores and predicted GPA. In particular, 14 of 42 subjects did so on the sense of humor tasks. Few produced non-monotonic responses with GPA (0) and mental concentration (4), suggesting that non-monotonicity reflected a response to the lower validity of the sense of humor test rather than random noise or confusion. Although not normatively valid, providing inconsistent responses is one way of showing little confidence in the sense of humor scores. The substantial regression observed with the sense of humor group in Figure 1a could be due in large part to these non-monotonic responses. Such response patterns lead to the inclusion of unusually high GPA's in the means associated with low percentiles and unusually low GPA's in the means associated with high percentiles. Figure 1b and Table 7 show the results of excluding all subjects with non-monotonic responses. The pattern was similar to that found with all subjects although it was considerably attenuated.

Kahneman and Tversky (1973) drew the subjects in their study from the same population (paid volunteers recruited through the University of Oregon student newspaper). As they report no culling of subjects, it may be that the slight regression reported in their between-subject design was due to the inclusion of some non-monotonic responses.

<u>3 Percentiles</u>. Figure 2a and Table 7 show results from all subjects who made GPA judgments for 3-percentile scores. Regression with the less valid scores is apparent. Again, several subjects (8 out of 44) had non-monotonic

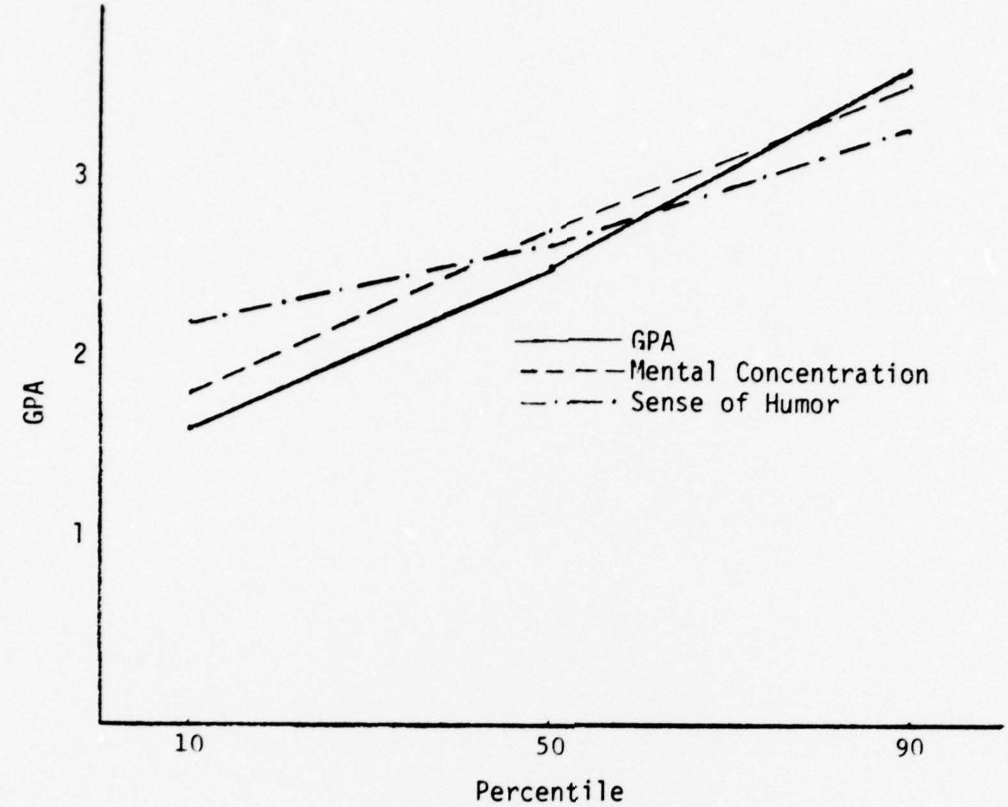


Figure 1a

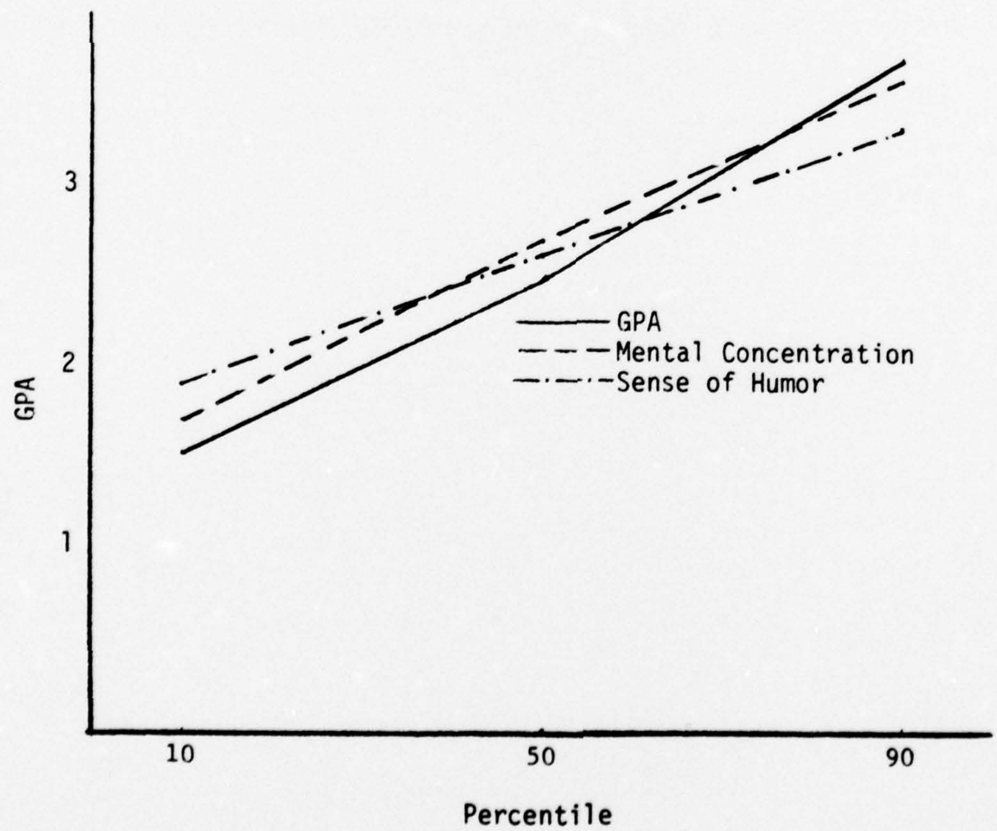


Figure 2a

Predictions of grade point average from 3 percentile scores on 3 variables. (a) All subjects (b) All subjects providing monotonic responses

responses to the sense of humor task, as did lesser numbers with GPA and mental concentration (2 each). Tables 6 and 7 and Figure 2b delete these subjects and reveal the same pattern of regression, slightly weakened. The majority of subjects exhibited reduced ranges with the less valid information.

In Table 6 the spread in GPA judgments induced by having to consider 11 percentiles is clearly visible. For each type of score, the mean range is greater by about 1.0 with 11 than with 3-percentile scores. The greatest mean range with 3 scores (GPA) is smaller than the smallest range with 11 scores (sense of humor).

Discussion

In the present within-subject design, subjects exhibited a sensitivity to validity information not apparent in Tversky and Kahneman's (1973) between-subject study. Whether judged by the ranges of GPA's or the proportion giving non-monotonic responses (2% with GPA; 7% with mental concentration; 27% with sense of humor), subjects responded differently when predicting on the basis of poorer quality information.

As in the studies of sensitivity to base-rate information, while subjects generally showed the right kind of sensitivity to the informational variable that was systematically varied (here validity), they do not seem to have been sufficiently sensitive. An accurate measure of the validity of each score is needed to ascertain the precise amount of regression needed. However, sense of humor scores

seem to be too close to the other scores, and while mental concentration judgments are regressed relative to GPA judgments, the difference is fairly small.

## 4. EXPERIMENT 3--SAMPLE SIZE

Kahneman and Tversky (1972) had subjects estimate sampling distributions of the percentage of boys among the N babies born in a certain region daily. For N = 1,000, their question read:

> On what percentage of days will the number of boys among 1,000 babies be as follows:
>
> Up to 50 boys
> 50 to 150 boys
> 150 to 250 boys
> . . . . . . . .
> 850 to 950 boys
> More than 950 boys
>
> Note that the categories include all possibilities, so your answers should add up to about 100%.

For N = 100, the categories were: Up to 5, 5-15, . . . , 85-95, More than 95 boys. For N = 10, each category contained a single outcome, e.g., 6 boys. Each group of subjects received only one value of N.

The startling result of their study was that virtually the same distributions were received with the very different values of N. Of course, it is much less probable to receive 5% (or 95%) boys in a sample of 1,000 than in a sample of 100. They interpreted this result in terms of the representativeness heuristic, according to which the likelihood of a sample is judged by the degree to which it represents the salient feature(s) of the population from which it is drawn. Since sample size is not a characteristic of populations, it is ignored in inferences regarding samples.

Experiment 3 converted their between-subject design to a within-subject design, with each subject assessing the likelihood of 3 sampling distributions, with N = 10, 100 and 1,000. Although this particular manipulation has not been attempted previously, Kahneman and Tversky (1972) and others (Bar-Hillel, Note 1) have demonstrated insensitivity to sample size within subjects. For example, Kahneman and Tversky (1972) asked subjects whether a hospital in which 45 babies were born daily or one in which 15 were born daily would have more days with more than 60% boys. Most of their subjects thought that the number of such days would be similar for the two hospitals. The remainder were about equally divided as to which would have more. Thus, there was less reason to expect the enforced sensitivity analysis to work here than in Experiments 1 and 2.

## Method

Subjects were asked to estimate the percentage of days on which up to 5%, 5-15%, 15-25%, etc. boys would be born in regions with 10, 100 and 1,000 babies born daily. The format quoted above was used for all questions. Half of the subjects considered N's of 10, 100, and 1,000 in that order; for the remainder, the order was reversed. Thirty-eight subjects were recruited as before.

## Results

As the order of presentation made no difference in the responses, the data from the two orders were combined. Inspection of the data revealed 4 subjects whose subjective sampling distributions either were not single peaked or whose peak was at an end category. Data from these subjects were eliminated.

Figure 3 shows, for the remaining 34 subjects, the median probabilities associated with each category for N = 10 and N = 1,000. They are remarkably similar. The distribution for N - 100 was also quite similar. Only three of the 38 subjects produced distributions for which the tails thickened and center flattened as N decreased, as sampling theory dictates.

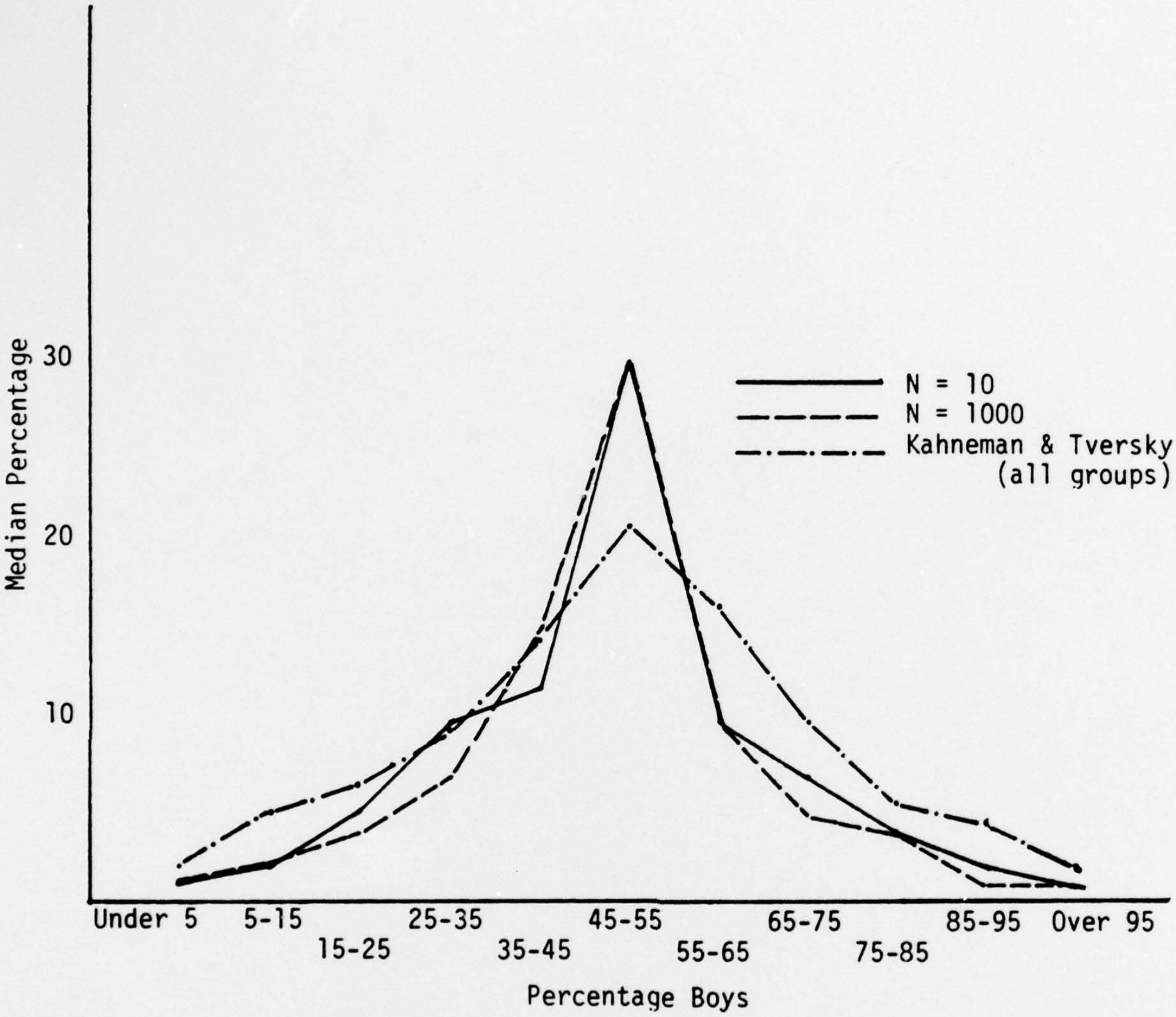


Figure 3
Estimated sampling distribution with N = 10, 100, 1,000.

## 5. GENERAL DISCUSSION

Three experiments forced subjects to consider the impact on their judgments of alternative values of three kinds of information found to be ignored in earlier experiments. For two kinds, base-rate information and predictive validity information, about two-thirds of subjects were influenced in the proper direction by changes in value. In neither case, however, were they sufficiently responsive. With the third kind of information, sample size, they showed no sensitivity at all. This mixture of results has both theoretical and practical implications.

On a theoretical level, the contrast between within-subject and between-subject designs suggests the need to temper previously made statements regarding the kinds of information that people neglect. Although effectively ignored when embedded in the context of other information, both base-rate and validity information elicit somewhat appropriate responses when varied systematically (for most subjects). Of course, there is an implicit demand not to respond the same way each time. But the differences in responses would not be properly ordered if subjects did not know (or were not able to figure out) the meanings of that information for their inferences. Experiment 3 showed that enforced sensitivity analyses do not guarantee more optimal responses.

Hammond and Summers (1972) have argued for a distinction between the judgmental strategies people wish to apply and those they actually apply. They attribute discrepancies between the desired and actual responses to a lack of cognitive control, the ability to implement desired

strategies. In this light, the present studies indicate that people have at their disposal judgmental strategies or heuristics that are more optimal than those demonstrated in earlier studies. Where this is the case, researchers interested in improving judgment might change the focus of their efforts from teaching people to use superior heuristics to inducing them to make more optimal use of the heuristics already at their disposal. Recommending that judges administer sensitivity analyses to themselves whenever they are required to combine several pieces of information might seem to be a generally useful strategy.

Before issuing a blanket recommendation, three issues must be confronted. The first is "When will it work?" If people apply an ineffective debiasing procedure they may succeed in increasing their confidence without improving their judgment, hardly a desirable combination. At the moment, it seems premature to predict effectiveness. Slovic and Fischhoff (1977) found that hindsight bias, the exaggerated tendency to view reported events as having appeared inevitable before they occurred, can be reduced by having people consider an alternate value of the event, that is, by relating how they would have explained the event had it turned out otherwise. Other biases have not yet been examined in a within-subject context. Such evidence seems the best way to determine whether people have alternative and more appropriate heuristics than those shown in between-subject studies.

The second issue is how to help those who respond in the right direction, but make too small an adjustment. Two possibilities suggested by Experiment 1 are to use very extreme values of the information in question, values that are

implausibly high or low from a substantive point of view, and to require a large number of judgments.

The third issue is what to do when people do not respond at all to subjective sensitivity analyses (or respond in the wrong direction). In such situations, it may be advisable to apply a correction factor to people's intuitive judgments or replace intuitions altogether by a formal rule (e.g., Bayes' Theorem) for combining information.

Finally, in situations where subjective sensitivity analysis is effective, some way is needed to highlight the similarities between structurally analogous situations (e.g., the cab and light bulb problems in Experiment 1) so as to induce some general learning. Kahneman and Tversky (1973) have argued pessimistically that even formal training in statistics will not guarantee sensitivity to non-intuitive effects like regression (due to low predictive validity) or increased variance (due to reduced sample size).

## 6. FOOTNOTES

1. GPA's at the University of Oregon range from 0.0 to 4.0. Although this range was not mentioned to subjects, all responses fell within it.

# 7. REFERENCE NOTE

1. Bar-Hillel, Maya. Additional Notes on Representativeness. Unpublished manuscript, Decision Research, August, 1977.

## 8. REFERENCES


Ajzen, I. Intuitive theories of events and the effects of base rate information on prediction. Journal of Personality and Social Psychology, 1977, 35: 303-314.

Bar-Hillel, M. The base-rate fallacy in probability judgments. Decision Research Report 77-4, 1977.

Fischhoff, B. The impact of information: Anticipated and actual. Decision Research Report 77-13, 1977.

Hammond, K.R. & Summers, D.A. Cognitive control. Psychological Review, 1972, 79: 58-67.

Kahneman, D. & Tversky, A. Subjective probability: A judgment of representativeness. Cognitive Psychology, 1972, 3: 430-454.

Kahneman, D. & Tversky, A. On the psychology of prediction. Psychological Review, 1973, 80: 237-251.

Lyon, D. & Slovic, P. Dominance of accuracy information and neglect of base rates in probability estimation. Acta Psychologica, 1973, 40: 287-298.

Nisbett, R.E. & Borgida, E. Attribution and the psychology of prediction. Journal of Personality and Social Psychology, 1977, 35: 613-624.

Selvidge, J. Experimental comparison of different methods for assessing the extremes of probability distributions by the fractile method. Management Science Report Series, Report 75-13. Boulder, Colorado: Graduate School of Business Administration, University of Colorado, 1975.

Slovic, P. & Fischhoff, B. On the psychology of experimental surprises. Journal of Experimental Psychology: Human Perception and Performance, 1977, 3: 544-551.

Tversky, A. & Kahneman, D. The belief in the "law of small numbers." Psychological Bulletin, 1971, 76: 105-110.

CONTRACT DISTRIBUTION LIST
(Unclassified Technical Reports)

| | |
|---|---|
| Director<br>Advanced Research Projects Agency<br>Attention: Program Management Office<br>1400 Wilson Boulevard<br>Arlington, Virginia 22209 | 2 copies |
| Office of Naval Research<br>Attention: Code 455<br>800 North Quincy Street<br>Arlington, Virginia 22217 | 3 copies |
| Defense Documentation Center<br>Attention: DDC-TC<br>Cameron Station<br>Alexandria, Virginia 22314 | 12 copies |
| DCASMA Baltimore Office<br>Attention: Mr. K. Gerasim<br>300 East Joppa Road<br>Towson, Maryland 21204 | 1 copy |
| Director<br>Naval Research Laboratory<br>Attention: Code 2627<br>Washington, D.C. 20375 | 6 copies |
| Decisions and Designs, Incorporated<br>8400 Westpark Drive, P.O. Box 907<br>McLean, Virginia 22101 | 10 copies |

SUPPLEMENTAL DISTRIBUTION LIST
(Unclassified Technical Reports)

Department of Defense

Director of Net Assessment
Office of the Secretary of Defense
Attention: MAJ Robert G. Gough, USAF
The Pentagon, Room 3A930
Washington, DC 20301

Assistant Director (Net Technical Assessment)
Office of the Deputy Director of Defense Research and Engineering (Test and Evaluation)
The Pentagon, Room 3C125
Washington, DC 20301

Director, Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, VA 22209

Director, Cybernetics Technology Office
Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, VA 22209

Director, ARPA Regional Office (Europe)
Headquarters, U.S. European Command
APO New York 09128

Director, ARPA Regional Office (Pacific)
Staff CINCPAC, Box 13
APO San Francisco 96610

Dr. Don Hirta
Defense Systems Management School
Building 202
Ft. Belvoir, VA 22060

Chairman, Department of Curriculum Development
National War College
Ft. McNair, 4th and P Streets, SW
Washington, DC 20319

Defense Intelligence School
Attention: Professor Douglas E. Hunter
Washington, DC 20374

Vice Director for Production
Management Office (Special Actions)
Defense Intelligence Agency
Room 1E863, The Pentagon
Washington, DC 20301

Command and Control Technical Center
Defense Communications Agency
Attention: Mr. John D. Hwang
Washington, DC 20301

Department of the Navy

Office of the Chief of Naval Operations (OP-951)
Washington, DC 20450

Office of Naval Research
Assistant Chief for Technology (Code 200)
800 N. Quincy Street
Arlington, VA 22217

Office of Naval Research (Code 230)
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research
Naval Analysis Programs (Code 431)
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research
Operations Research Programs (Code 434)
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research
Information Systems Program (Code 437)
800 North Quincy Street
Arlington, VA 22217

Director, ONR Branch Office
Attention: Dr. Charles Davis
536 South Clark Street
Chicago, IL 60605

Director, ONR Branch Office
Attention: Dr. J. Lester
495 Summer Street
Boston, MA 02210

Director, ONR Branch Office
Attention: Dr. E. Gloye
1030 East Green Street
Pasadena, CA 91106

Director, ONR Branch Office
Attention: Mr. R. Lawson
1030 East Green Street
Pasadena, CA 91106

Office of Naval Research
Scientific Liaison Group
Attention: Dr. M. Bertin
American Embassy - Room A-407
APO San Francisco 96503

Dr. A. L. Slafkosky
Scientific Advisor
Commandant of the Marine Corps (Code RD-1)
Washington, DC 20380

Headquarters, Naval Material Command
(Code 0331)
Attention: Dr. Heber G. Moore
Washington, DC 20360

Dean of Research Administration
Naval Postgraduate School
Attention: Patrick C. Parker
Monterey, CA 93940

Superintendent
Naval Postgraduate School
Attention: R. J. Roland, (Code 52R1)
$C^3$ Curriculum
Monterey, CA 93940

Naval Personnel Research and Development
Center (Code 305)
Attention: LCDR O'Bar
San Diego, CA 92152

Navy Personnel Research and Development
Center
Manned Systems Design (Code 311)
Attention: Dr. Fred Muckler
San Diego, CA 92152

Naval Training Equipment Center
Human Factors Department (Code N215)
Orlando, FL 32813

Naval Training Equipment Center
Training Analysis and Evaluation Group
(Code N-00T)
Attention: Dr. Alfred F. Smode
Orlando, FL 32813

Director, Center for Advanced Research
Naval War College
Attention: Professor C. Lewis
Newport, RI 02840

Naval Research Laboratory
Communications Sciences Division (Code 5403)
Attention: Dr. John Shore
Washington, DC 20375

Dean of the Academic Departments
U.S. Naval Academy
Annapolis, MD 21402

Chief, Intelligence Division
Marine Corps Development Center
Quantico, VA 22134

<u>Department of the Army</u>

Deputy Under Secretary of the Army
(Operations Research)
The Pentagon, Room 2E621
Washington, DC 20310

Director, Army Library
Army Studies (ASDIRS)
The Pentagon, Room 1A534
Washington, DC 20310

U.S. Army Research Institute
Organizations and Systems Research Laboratory
Attention: Dr. Edgar M. Johnson
5001 Eisenhower Avenue
Alexandria, VA 22333

Director, Organizations and Systems
Research Laboratory
U.S. Army Institute for the Behavioral
and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

Technical Director, U.S. Army Concepts
Analysis Agency
8120 Woodmont Avenue
Bethesda, MD 20014

Director, Strategic Studies Institute
U.S. Army Combat Developments Command
Carlisle Barracks, PA 17013

Commandant, Army Logistics Management Center
Attention: DRXMC-LS-SCAD (ORSA)
Ft. Lee, VA 23801

Department of Engineering
United States Military Academy
Attention: COL A. F. Grum
West Point, NY 10996

Marine Corps Representative
U.S. Army War College
Carlisle Barracks, PA 17013

Chief, Studies and Analysis Office
Headquarters, Army Training and Doctrine
Command
Ft. Monroe, VA 23351

Commander, U.S. Army Research Office
(Durham)
Box CM, Duke Station
Durham, NC 27706

Department of the Air Force

Assistant for Requirements Development
and Acquisition Programs
Office of the Deputy Chief of Staff for
Research and Development
The Pentagon, Room 4C331
Washington, DC 20330

Air Force Office of Scientific Research
Life Sciences Directorate
Building 410, Bolling AFB
Washington, DC 20332

Commandant, Air University
Maxwell AFB, AL 36112

Chief, Systems Effectiveness Branch
Human Engineering Division
Attention: Dr. Donald A. Topmiller
Wright-Patterson AFB, OH 45433

Deputy Chief of Staff, Plans, and
Operations
Directorate of Concepts (AR/XOCCC)
Attention: Major R. Linhard
The Pentagon, Room 4D 1047
Washington, DC 20330

Director, Advanced Systems Division
(AFHRL/AS)
Attention: Dr. Gordon Eckstrand
Wright-Patterson AFB, OH 45433

Commander, Rome Air Development Center
Attention: Mr. John Atkinson
Griffis AFB
Rome, NY 13440

IRD, Rome Air Development Center
Attention: Mr. Frederic A. Dion
Griffis AFB
Rome, NY 13440

HQS Tactical Air Command
Attention: LTCOL David Dianich
Langley AFB, VA 23665

## Other Government Agencies

Chief, Strategic Evaluation Center
Central Intelligence Agency
Headquarters, Room 2G24
Washington, DC 20505

Director, Center for the Study of Intelligence
Central Intelligence Agency
Attention: Mr. Dean Moor
Washington, DC 20505

Mr. Richard Heuer
Methods & Forecasting Division
Office of Regional and Political Analysis
Central Intelligence Agency
Washington, DC 20505

Office of Life Sciences
Headquarters, National Aeronautics and Space Administration
Attention: Dr. Stanley Deutsch
600 Independence Avenue
Washington, DC 20546

## Other Institutions

Department of Psychology
The Johns Hopkins University
Attention: Dr. Alphonse Chapanis
Charles and 34th Streets
Baltimore, MD 21218

Institute for Defense Analyses
Attention: Dr. Jesse Orlansky
400 Army Navy Drive
Arlington, VA 22202

Director, Social Science Research Institute
University of Southern California
Attention: Dr. Ward Edwards
Los Angeles, CA 90007

Perceptronics, Incorporated
Attention: Dr. Amos Freedy
6271 Variel Avenue
Woodland Hills, CA 91364 (10 copies)

Stanford University
Attention: Dr. R. A. Howard
Stanford, CA 94305

Director, Applied Psychology Unit
Medical Research Council
Attention: Dr. A. D. Baddeley
15 Chaucer Road
Cambridge, CB 2EF
England

Department of Psychology
Brunel University
Attention: Dr. Lawrence D. Phillips
Uxbridge, Middlesex UB8 3PH
England

Decision Analysis Group
Stanford Research Institute
Attention: Dr. Miley W. Merkhofer
Menlo Park, CA 94025

Decision Research
1201 Oak Street
Eugene, OR 97401 (10 copies)

Department of Psychology
University of Washington
Attention: Dr. Lee Roy Beach
Seattle, WA 98195

Department of Electrical and Computer Engineering
University of Michigan
Attention: Professor Kan Chen
Ann Arbor, MI 94135

Department of Government and Politics
University of Maryland
Attention: Dr. Davis B. Bobrow
College Park, MD 20747

Department of Psychology
Hebrew University
Attention: Dr. Amos Tversky
Jerusalem, Israel

Dr. Andrew P. Sage
School of Engineering and Applied Science
University of Virginia
Charlottesville, VA 22901

Professor Raymond Tanter
Political Science Department
The University of Michigan
Ann Arbor, MI 48109

Professor Howard Raiffa
Morgan 302
Harvard Business School
Harvard University
Cambridge, MA 02163

Department of Psychology
University of Oklahoma
Attention: Dr. Charles Gettys
455 West Lindsey
Dale Hall Tower
Norman, OK 73069

Institute of Behavioral Science #3
University of Colorado
Attention: Dr. Kenneth Hammond
Room 201
Boulder, Colorado 80309

unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

| REPORT DOCUMENTATION PAGE | | READ INSTRUCTIONS BEFORE COMPLETING FORM |
|---|---|---|
| 1. REPORT NUMBER<br>PTR-1060-78-7 | 2. GOVT ACCESSION NO. | 3. RECIPIENT'S CATALOG NUMBER |
| 4. TITLE (and Subtitle)<br>Improving Intuitive Judgments by Subjective Sensitivity Analysis | | 5. TYPE OF REPORT & PERIOD COVERED<br>Technical Rep[redacted]t |
| | | 6. PERFORMING ORG. REPORT NUMBER |
| 7. AUTHOR(s)<br>Baruch Fischhoff, Paul Slovic [redacted] Sarah Lichtenstein | | 8. CONTRACT OR GRANT NUMBER(s)<br>N00014-78-C-0100 |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS<br>Decision Research, A Branch of Perceptronics<br>1201 Oak Street<br>Eugene, Oregon 97401 | | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS<br>ARPA Order 3469 |
| 11. CONTROLLING OFFICE NAME AND ADDRESS<br>Defense Advanced Research Projects Agency<br>1400 Wilson Blvd.<br>Arlington, Virginia 22217 | | 12. REPORT DATE<br>Jul[redacted]978 |
| | | 13. NUMBER OF PAGES<br>52 |
| 14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)<br>Office of Naval Research, Code 455<br>800 North Quincy Street<br>Arlington, Virginia 22217 | | 15. SECURITY CLASS. (of this report)<br>unclassified |
| | | 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE |

16. DISTRIBUTION STATEMENT (of this Report)

approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

none

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

subjective sensitivity analysis
intuitive judgments
judgmental bias
decision making

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Previous studies have concluded that when making inferences, people tend to ignore various kinds of normatively important information (e.g., samples sizes, base rates). The basis for such conclusions has typically been the similarity of responses across groups presented with widely discrepant values of that information. The experiments reported here convert earlier, between-subject designs to within-subject formats. Each subject made several judgments as one kind of information varied across a range of possible values. When

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

390 664

information regarding base rates and predictive validity was varied, roughly two-thirds of the subjects changed their judgments in directions dictated by normative considerations. There were no appreciable shifts in response to changes in sample size. Apparently, people know (or can figure out) somewhat more than what they have been given credit for in the past.